**ПРОБЛЕМЫ ПЕРЕДАЧИ ИНФОРМАЦИИ**

**Том 57** **2021** **Вып. 1**

УДК 621.391 : 519.72

 **В.В. Прелов**

# $f$-ДИВЕРГЕНЦИЯ И СКЛЕИВАНИЕ ВЕРОЯТНОСТНЫХ РАСПРЕДЕЛЕНИЙ[1]

Рассматривается задача о нахождении минимальных и максимальных значений $f$-дивергенции дискретных распределений вероятностей $P$ и $Q$ при условии, что заданы одно из этих распределений и величина их склеивания. Для минимума $f$-дивергенции при указанных условиях получено явное выражение, а для ее максимума – точное выражение, которое в общем случае не является явным, но для многих частных случаев позволяет выписать как явные формулы, так и простые верхние границы, являющиеся в некоторых случаях оптимальными. Подобные явные формулы и верхние границы получены для дивергенции Кульбака – Лейблера и $\chi^2$-дивергенции, являющихся важнейшими частными случаями $f$-дивергенции.

*Ключевые слова*: $f$-дивергенция, дивергенция Кульбака – Лейблера, $\chi^2$-дивергенция, склеивание дискретных распределений вероятностей.

**DOI:** 10.31857/S0555292321010034

## § 1. Введение и формулировки основных результатов

Пусть $P = \{p_i,\ i \in \mathcal{N}\}$ и $Q = \{q_i,\ i \in \mathcal{N}\}$ – дискретные распределения вероятностей со значениями в конечном множестве $\mathcal{N} = \{1, 2, \ldots, n\}$. Напомним, что $f$-дивергенция $D_f(P \,\|\, Q)$ распределения $P$ относительно $Q$ определяется равенством (см. [1, 2])

$$D_f(P \,\|\, Q) = \sum_{i \in \mathcal{N}} q_i f\left(\frac{p_i}{q_i}\right), \tag{1}$$

где $f\colon (0, \infty) \to \mathbb{R}$ – выпуклая функция, такая что $f(1) = 0$ (в дальнейшем будем всегда считать, что $f(\cdot)$ – дважды дифференцируемая функция, такая что $f''(x) > 0$, $x \neq 1$). При этом всегда по определению предполагается, что

$$0 \cdot f\left(\frac{0}{0}\right) = 0, \quad f(0) = \lim_{u \downarrow 0} f(u), \quad 0 \cdot f\left(\frac{a}{0}\right) = \lim_{\varepsilon \downarrow 0} \varepsilon f\left(\frac{a}{\varepsilon}\right) = a \lim_{t \to \infty} \frac{f(t)}{t},$$

где $a \neq 0$. Частными случаями $f$-дивергенции являются многие известные меры различия между распределениями вероятностей, используемые в теории информации, теории вероятностей и математической статистике. Наиболее важными примерами $f$-дивергенций являются дивергенция Кульбака – Лейблера (или просто дивергенция)

$$D(P \,\|\, Q) = \sum_{i \in \mathcal{N}} p_i \log \frac{p_i}{q_i} = D_f(P \,\|\, Q)$$

[1] Работа выполнена при частичной финансовой поддержке Российского фонда фундаментальных исследований (номер проекта 19-01-00364).

при $f(t) = t\log t$ и $\chi^2$-дивергенция

$$\chi^2(P\,\|\,Q) = \sum_{i\in\mathcal{N}} \frac{(p_i - q_i)^2}{q_i} = D_f(P\,\|\,Q)$$

при $f(t) = (t-1)^2$, а также вариационное расстояние, дивергенция Хеллингера и др. (см., например, [3, 4]).

Напомним также, что $\alpha$-склеиванием дискретных распределений вероятностей $P = \{p_i,\ i \in \mathcal{N}\}$ и $Q = \{q_i,\ i \in \mathcal{N}\}$ называется совместное распределение $P_{XY}$ случайных величин $X$ и $Y$ со значениями в множестве $\mathcal{N}$ и маргинальными распределениями $P_X = P$ и $P_Y = Q$, такое что $\Pr\{X = Y\} = \alpha$ (см. [5]). В дальнейшем величину склеивания распределений $P$ и $Q$ будем обозначать через $s(P, Q)$.

В работе [6] рассматривалась задача о нахождении минимальных и максимальных значений дивергенции Реньи $D_\lambda(P\,\|\,Q)$ при условии, что заданы одно из распределений $P$ или $Q$ и величина их склеивания $s(P, Q)$. В настоящей статье рассматривается аналогичная задача о нахождении минимальных и максимальных значений для произвольной $f$-дивергенции дискретных распределений вероятностей $P$ и $Q$, а также ее важнейших частных случаев – дивергенции Кульбака – Лейблера и $\chi^2$-дивергенции, которые, например, используются при получении границ для коэффициентов сжатия $f$-дивергенций [4]. Отметим, что задача о нахождении экстремальных значений $f$-дивергенции $D_f(P\,\|\,Q)$, когда вместо условия $\alpha$-склеивания накладывалось условие на вариационное расстояние между $P$ и $Q$, рассматривалась в [7–9].

Для формулировки полученных результатов введем необходимые определения и обозначения. Для заданных распределения вероятностей $P = \{p_i,\ i \in \mathcal{N}\}$, действительного числа $\alpha$, $0 \leqslant \alpha \leqslant 1$, и выпуклой функции $f(\cdot)$, задающей $f$-дивергенцию $D_f(P\,\|\,Q)$, определим величину $D_f^{\min}(P, \alpha)$ равенством

$$D_f^{\min}(P, \alpha) = \min_{Q:\, s(P,Q)=\alpha} D_f(P\,\|\,Q), \tag{2}$$

где минимум берется по всевозможным распределениям $Q = \{q_i,\ i \in \mathcal{N}\}$, для которых существует их $\alpha$-склеивание с распределением $P$. Аналогично определяется и величина $D_f^{\min}(Q, \alpha)$, если задано распределение $Q = \{q_i,\ i \in \mathcal{N}\}$, т.е.

$$D_f^{\min}(Q, \alpha) = \min_{P:\, s(P,Q)=\alpha} D_f(P\,\|\,Q). \tag{3}$$

В случае, когда $P = \{p, 1-p\}$ и $Q = \{q, 1-q\}$, вместо $D_f(P\,\|\,Q)$ будем использовать обозначение $d_f(p\,\|\,q)$, т.е.

$$d_f(p\,\|\,q) = qf\left(\frac{p}{q}\right) + (1-q)f\left(\frac{1-p}{1-q}\right). \tag{4}$$

Заметим, что из свойств функции $f(\cdot)$ (выпуклости и равенства $f(1) = 0$) следует, что $d_f(p\,\|\,q)$ является выпуклой неотрицательной функцией как параметра $p$, так и параметра $q$, причем $d_f(p\,\|\,q) = 0$ при $p = q$.

Теорема 1. *Справедливы следующие равенства*:

$$D_f^{\min}(P, \alpha) = \begin{cases} 0, & \text{если } p_{\max} \leqslant \dfrac{1}{2} + \dfrac{\alpha}{2}, \\ d_f(p_{\max}\,\|\,1 - p_{\max} + \alpha), & \text{если } p_{\max} \geqslant \dfrac{1}{2} + \dfrac{\alpha}{2}, \end{cases} \tag{5}$$

$$D_f^{\min}(Q,\alpha)=\begin{cases}0, & \text{если } q_{\max}\leqslant \dfrac{1}{2}+\dfrac{\alpha}{2},\\ d_f(1-q_{\max}+\alpha\,\|\,q_{\max}), & \text{если } q_{\max}\geqslant \dfrac{1}{2}+\dfrac{\alpha}{2},\end{cases} \tag{6}$$

*где* $p_{\max}=\max\limits_{i\in\mathcal N} p_i$ *и* $q_{\max}=\max\limits_{i\in\mathcal N} q_i$.

Доказательства этой и нижеследующих теорем приведены в § 2. Отметим, что частный случай теоремы 1, когда $f(t)=t\log t$ или $f(t)=-\log t$, т.е. для дивергенции Кульбака – Лейблера, был доказан в [4].

Для формулировки следующей теоремы введем еще несколько определений. Для заданных распределения вероятностей $P=\{p_i,\ i\in\mathcal N\}$, действительного числа $\alpha$, $0\leqslant\alpha\leqslant 1$, и выпуклой функции $f(\cdot)$, задающей $f$-дивергенцию $D_f(P\,\|\,Q)$, определим величину $D_f^{\max}(P,\alpha)$ равенством

$$D_f^{\max}(P,\alpha)=\max_{Q\colon s(P,Q)=\alpha} D_f(P\,\|\,Q), \tag{7}$$

где максимум берется по всевозможным распределениям $Q=\{q_i,\ i\in\mathcal N\}$, для которых существует их $\alpha$-склеивание с распределением $P$. Аналогично определяется и величина $D_f^{\max}(Q,\alpha)$, если задано распределение $Q=\{q_i,\ i\in\mathcal N\}$, т.е.

$$D_f^{\max}(Q,\alpha)=\max_{P\colon s(P,Q)=\alpha} D_f(P\,\|\,Q). \tag{8}$$

Всякое равенство

$$\alpha=\sum_{i\in I} p_i+\beta,\quad \text{где } I\subseteq\mathcal N, \tag{9}$$

назовем *допустимым* $(P,I)$*-представлением* $\alpha$, если либо $\beta=0$, либо существует индекс $j\in\mathcal N\setminus I$, такой что $0<\beta<p_j$. Аналогично определяется *допустимое* $(Q,I)$*-представление* $\alpha$, если задано распределение $Q=\{q_i,\ i\in\mathcal N\}$ и параметр $\alpha$, $0\leqslant\alpha\leqslant 1$.

Всякое $\alpha$-склеивание заданного распределения вероятностей $P=\{p_i,\ i\in\mathcal N\}$ с некоторым распределением $Q=\{q_i,\ i\in\mathcal N\}$ задается с помощью квадратной матрицы $M=\|p_{ij}\|_{i,j=1}^n$ с неотрицательными элементами $p_{ij}$, такой что $\sum\limits_{j=1}^n p_{ij}=p_i$ для всех $i\in\mathcal N$, $\sum\limits_{i=1}^n p_{ij}=q_j$ для всех $j\in\mathcal N$ и $\sum\limits_{i=1}^n p_{ii}=\alpha$ . В этом случае положим $D_f(M)=D_f(P\,\|\,Q)$. Аналогично (с заменой в предыдущем определении распределения $P$ на $Q$ и наоборот) задается $\alpha$-склеивание данного распределения вероятностей $Q=\{q_i,\ i\in\mathcal N\}$ с некоторым распределением $P=\{p_i,\ i\in\mathcal N\}$.

Каждому допустимому $(P,I)$-представлению $\alpha$ сопоставим множество $\mathcal M(P,I)$ матриц $M=\|p_{ij}\|_{i,j=1}^n$, осуществляющих $\alpha$-склеивание заданного распределения вероятностей $P=\{p_i,\ i\in\mathcal N\}$ с некоторым распределением $Q=\{q_i,\ i\in\mathcal N\}$ и обладающих следующим свойством: на (главной) диагонали каждой такой матрицы стоят числа $p_i$ и $\beta$, входящие в данное допустимое $(P,I)$-представление $\alpha$, а все остальные ненулевые элементы матрицы находятся в некотором столбце (будем называть такой столбец *главным*) и, возможно, лишь один ненулевой элемент находится вне диагонали и этого главного столбца. Аналогично, каждому допустимому $(Q,I)$-представлению $\alpha$ сопоставляется множество $\mathcal M(Q,I)$ матриц $M=\|p_{ij}\|_{i,j=1}^n$, осуществляющих $\alpha$-склеивание заданного распределения вероятностей $Q=\{q_i,\ i\in\mathcal N\}$ с некоторым распределением $P=\{p_i,\ i\in\mathcal N\}$.

Т е о р е м а 2. *Для любого распределения вероятностей*

$$P=\{p_i,\ i\in\mathcal N\},\quad p_1\geqslant p_2\geqslant\ldots\geqslant p_n>0,$$

*и любого* $\alpha$, $0 \leqslant \alpha \leqslant 1$, *справедливо равенство*

$$D_f^{\max}(P,\alpha) = \max_{(P,I)} \max_{M \in \mathcal{M}(P,I)} D_f(M), \tag{10}$$

*где первый максимум в правой части* (10) *берется по всем допустимым* $(P,I)$*-представлениям* $\alpha$. *В частности,*

$$D_f^{\max}(P,\alpha) = \begin{cases} \infty, & \text{если } \alpha \leqslant 1-p_n \text{ и } f^* = \infty, \\ K_P, & \text{если } \alpha > 1-p_n, \end{cases} \tag{11}$$

*где* $f^* = \lim\limits_{t\to\infty} \dfrac{f(t)}{t}$, *а*

$$K_P = \max\biggl\{(\alpha-1+p_n)f\biggl(\frac{p_n}{\alpha-1+p_n}\biggr) + (1-\alpha+p_{n-1})f\biggl(\frac{p_{n-1}}{1-\alpha+p_{n-1}}\biggr), \\ (\alpha-1+p_{n-1})f\biggl(\frac{p_{n-1}}{\alpha-1+p_{n-1}}\biggr) + (1-\alpha+p_n)f\biggl(\frac{p_n}{1-\alpha+p_n}\biggr)\biggr\}. \tag{12}$$

Во многом аналогичная теорема справедлива и для величины $D_f^{\max}(Q,\alpha)$, определенной в (8).

Теорема 3. *Для любого распределения вероятностей*

$$Q = \{q_i,\ i \in \mathcal{N}\}, \quad q_1 \geqslant q_2 \geqslant \ldots \geqslant q_n > 0,$$

*и любого* $\alpha$, $0 \leqslant \alpha \leqslant 1$, *справедливо равенство*

$$D_f^{\max}(Q,\alpha) = \max_{(Q,I)} \max_{M \in \mathcal{M}(Q,I)} D_f(M), \tag{13}$$

*где первый максимум в правой части* (13) *берется по всем допустимым* $(Q,I)$*-представлениям* $\alpha$. *В частности,*

$$D_f^{\max}(Q,\alpha) = \begin{cases} \infty, & \text{если } \alpha \leqslant 1-q_n \text{ и } f(0) = \infty, \\ K_Q, & \text{если } \alpha > 1-q_n, \end{cases} \tag{14}$$

*где*

$$K_Q = \max\biggl\{q_n f\biggl(\frac{\alpha-1+q_n}{q_n}\biggr) + q_{n-1} f\biggl(\frac{1-\alpha+q_{n-1}}{q_{n-1}}\biggr), \\ q_{n-1} f\biggl(\frac{\alpha-1+q_{n-1}}{q_{n-1}}\biggr) + q_n f\biggl(\frac{1-\alpha+q_n}{q_n}\biggr)\biggr\}. \tag{15}$$

Как видно из формулировок теорем 2 и 3, формулы (10) и (13) не позволяют для общего случая $f$-дивергенции выписывать явные выражения для $D_f^{\max}(P,\alpha)$ при $\alpha \leqslant 1-p_n$ и $f^* < \infty$ и для $D_f^{\max}(Q,\alpha)$ при $\alpha \leqslant 1-q_n$ и $f(0) < \infty$. Однако для многих конкретных $f$-дивергенций эти формулы позволяют получить как хорошие явные верхние границы для $D_f^{\max}(P,\alpha)$ и $D_f^{\max}(Q,\alpha)$ (которые в некоторых случаях являются оптимальными), так и явные выражения для них при малых значениях $\alpha$. Ниже мы покажем это на примерах дивергенции Кульбака–Лейблера и $\chi^2$-дивергенции.

Обозначим

$$D^{\max}(P,\alpha) = \max_{Q:\, s(P,Q)=\alpha} D(P \,\|\, Q) = D_f^{\max}(P,\alpha) \quad \text{при } f(t) = t\log t, \tag{16}$$

$$D^{\max}(Q,\alpha) = \max_{P:\, s(P,Q)=\alpha} D(P\,\|\,Q) = D_f^{\max}(Q,\alpha) \quad \text{при } f(t) = t\log t, \tag{17}$$

$$\chi^2_{\max}(P,\alpha) = \max_{Q:\, s(P,Q)=\alpha} \chi^2(P\,\|\,Q) = D_f^{\max}(P,\alpha) \quad \text{при } f(t) = (t-1)^2, \tag{18}$$

$$\chi^2_{\max}(Q,\alpha) = \max_{P:\, s(P,Q)=\alpha} \chi^2(P\,\|\,Q) = D_f^{\max}(Q,\alpha) \quad \text{при } f(t) = (t-1)^2, \tag{19}$$

где $D_f^{\max}(P,\alpha)$ и $D_f^{\max}(Q,\alpha)$ определены в (7) и (8) соответственно.

Теорема 4. *Для величин* $D^{\max}(P,\alpha)$ *и* $D^{\max}(Q,\alpha)$, *определенных в* (16) *и* (17), *справедливы следующие утверждения*:

- *Если заданы распределение вероятностей* $P = \{p_i, i \in \mathcal{N}\}$, $p_1 \geqslant p_2 \geqslant \ldots \geqslant p_n > 0$, *и число* $\alpha$, $0 \leqslant \alpha \leqslant 1$, *то*

$$D^{\max}(P,\alpha) = \\ = \begin{cases} \infty, & \text{если } \alpha \leqslant 1 - p_n, \\ p_n \log \dfrac{p_n}{p_n - 1 + \alpha} + p_{n-1} \log \dfrac{p_{n-1}}{p_{n-1} + 1 - \alpha}, & \text{если } \alpha > 1 - p_n; \end{cases} \tag{20}$$

- *Если заданы распределение вероятностей* $Q = \{q_i, i \in \mathcal{N}\}$, $q_1 \geqslant q_2 \geqslant \ldots \geqslant q_n > 0$, *и число* $\alpha$, $0 \leqslant \alpha \leqslant 1$, *то*

$$D^{\max}(Q,\alpha) = (1 + \alpha - q_n) \log \frac{1 + \alpha - q_n}{q_n} + (q_n - \alpha) \log \frac{q_n - \alpha}{q_{n-1}}, \tag{21}$$

*если* $\alpha \leqslant q_n$, *и*

$$D^{\max}(Q,\alpha) = (q_n - 1 + \alpha) \log \frac{q_n - 1 + \alpha}{q_n} + (1 - \alpha + q_{n-1}) \log \frac{1 - \alpha + q_{n-1}}{q_{n-1}}, \tag{22}$$

*если* $\alpha > 1 - q_n$;

- *Для всех* $\alpha$, $q_n \leqslant \alpha \leqslant 1 - q_n$, *справедлива верхняя граница*

$$D^{\max}(Q,\alpha) \leqslant (1 - \alpha + q_n) \log \frac{1 - \alpha + q_n}{q_n}, \tag{23}$$

*причем эта верхняя граница достигается, т.е. в* (23) *имеет место знак равенства, если* $\alpha = \sum\limits_{i=1}^{n-1} a_i q_i + q_n$ *при некоторых* $a_i \in \{0,1\}$.

Из этой теоремы можно также вывести следствие для величин $D^{\max}(p_{\min},\alpha)$ и $D^{\max}(q_{\min},\alpha)$, определяемых равенствами

$$D^{\max}(p_{\min},\alpha) = \max_{(P,Q):\, s(P,Q)=\alpha,\, \min\limits_{i\in\mathcal{N}} p_i = p_{\min}} D(P\,\|\,Q), \tag{24}$$

$$D^{\max}(q_{\min},\alpha) = \max_{(P,Q):\, s(P,Q)=\alpha,\, \min\limits_{i\in\mathcal{N}} q_i = q_{\min}} D(P\,\|\,Q), \tag{25}$$

где максимумы в (24), (25) берутся по всевозможным распределениям $P = \{p_i, i \in \mathcal{N}\}$ и $Q = \{q_i,\ i \in \mathcal{N}\}$ с заданными параметрами $p_{\min} > 0$ в (24) и $q_{\min} > 0$ в (25), таким что $s(P,Q) = \alpha$.

Следствие 1. *Для величин* $D^{\max}(p_{\min},\alpha)$ *и* $D^{\max}(q_{\min},\alpha)$, *определенных в* (24) *и* (25), *в случае* $|\mathcal{N}| = n \geqslant 3$ *справедливы следующие утверждения*:

- *Для всех* $p_{\min} > 0$ *и* $\alpha$, $0 \leqslant \alpha \leqslant 1$, *справедливо равенство*

$$D^{\max}(p_{\min}, \alpha) = \begin{cases} \infty, & \text{если } \alpha \leqslant 1 - p_{\min}, \\ p_{\min} \log \dfrac{p_{\min}^2}{p_{\min}^2 - (1-\alpha)^2}, & \text{если } \alpha > 1 - p_{\min}; \end{cases} \tag{26}$$

- *Для всех* $q_{\min} > 0$ *и* $\alpha$, *таких что* $0 \leqslant \alpha \leqslant q_{\min}$ *или* $1 - q_{\min} \leqslant \alpha \leqslant 1$, *справедливы равенства*

$$D^{\max}(q_{\min}, \alpha) = \log \frac{1}{q_{\min}} - h(1 + \alpha - q_{\min}), \quad \text{если } \alpha \leqslant q_{\min}, \tag{27}$$

$$D^{\max}(q_{\min}, \alpha) = 2q_{\min} \left[ \log 2 - h\left( \frac{1 - \alpha + q_{\min}}{2q_{\min}} \right) \right], \quad \text{если } \alpha \geqslant 1 - q_{\min}, \tag{28}$$

*где* $h(x) = -x \log x - (1-x)\log(1-x)$, $0 \leqslant x \leqslant 1$;

- *Для всех* $q_{\min} > 0$ *и* $\alpha$, $q_{\min} < \alpha < 1 - q_{\min}$, *справедлива верхняя граница*

$$D^{\max}(q_{\min}, \alpha) \leqslant (1 - \alpha + q_{\min}) \log \frac{1 - \alpha + q_{\min}}{q_{\min}}, \tag{29}$$

*причем эта верхняя граница достигается, т.е. в* (29) *имеет место знак равенства, если* $2q_{\min} \leqslant \alpha < 1 - q_{\min}$ *и* $q_{\min} \leqslant \dfrac{1}{n+1}$, *а также если* $q_{\min} \leqslant \dfrac{1}{n}$ *и* $\alpha = kq_{\min}$, *где* $k$ *– любое целое, такое что* $2 \leqslant k \leqslant n - 1$.

Доказательство. Равенство (26) является прямым следствием формулы (20), так как, с одной стороны,

$$p_{n-1} \log \frac{p_{n-1}}{p_{n-1} + 1 - \alpha} \leqslant p_n \log \frac{p_n}{p_n + 1 - \alpha},$$

а с другой стороны, если $|\mathcal{N}| \geqslant 3$, то всегда существует распределение вероятностей $P = \{p_i,\ i \in \mathcal{N}\}$, такое что $p_{n-1} = p_n = p_{\min}$. Аналогично доказывается, что и равенства (27), (28) являются прямыми следствиями соответствующих равенств (21), (22).

Наконец, достижение равенства в верхней границе (29) при сформулированных там условиях также следует из утверждения теоремы 4 о достижении верхней границы (23). Действительно, нетрудно предъявить соответствующее распределение вероятностей $Q = \{q_i,\ i \in \mathcal{N}\}$, зависящее от значения параметра $\alpha$, для которого имеет место равенство $\alpha = \sum\limits_{i=1}^{n-1} a_i q_i + q_n$ при некоторых $a_i \in \{0, 1\}$. А именно, если $kq_{\min} < \alpha < (k+1)q_{\min}$, где $k = 2, 3, \ldots, n-2$, или $(n-1)q_{\min} < \alpha < 1 - q_{\min}$, то очевидно, что $\alpha = \sum\limits_{i=1}^{k-1} q_i + q_n$ для распределения $Q = \{q_i,\ i \in \mathcal{N}\}$, компоненты $q_i$ которого задаются равенствами

$$q_i = \begin{cases} q_{\min} & \text{при} \quad i = 1, 2, \ldots, k-2, \\ \alpha - (k-1)q_{\min} & \text{при } i = k-1, \\ \dfrac{1-\alpha}{n-k} & \text{при } i = k, k+1, \ldots, n-1, \\ q_{\min} & \text{при } i = n, \end{cases}$$

и обладают тем свойством, что все $q_i \leqslant q_{\min}$, если $q_{\min} \leqslant \dfrac{1}{n+1}$.

Если же $\alpha = kq_{\min}$, где $k = 2, 3, \dots, n-1$, то снова очевидно, что $\alpha = \sum\limits_{i=1}^{k-1} q_i + q_n$ для распределения $Q = \{q_i,\ i \in \mathcal{N}\}$, компоненты $q_i$ которого задаются равенствами

$$q_i = \begin{cases} q_{\min} & \text{при } i = 1, 2, \dots, n-1, \\ 1 - (n-1)q_{\min} & \text{при } i = n, \end{cases}$$

и при этом все эти $q_i \leqslant q_{\min}$, если $q_{\min} \leqslant \dfrac{1}{n}$. ▲

Т е о р е м а 5. *Для величин* $\chi^2_{\max}(P, \alpha)$ *и* $\chi^2_{\max}(Q, \alpha)$, *определенных в* (18) *и* (19), *справедливы следующие утверждения*:

- *Если заданы распределение вероятностей* $P = \{p_i, i \in \mathcal{N}\}$, $p_1 \geqslant p_2 \geqslant \ldots \geqslant p_n > 0$, *и число* $\alpha$, $0 \leqslant \alpha \leqslant 1$, *то*

$$\chi^2_{\max}(P, \alpha) = \begin{cases} \infty, & \textit{если } \alpha \leqslant 1 - p_n, \\ \dfrac{(1-\alpha)^2}{\alpha + p_n - 1} + \dfrac{(1-\alpha)^2}{1 + p_{n-1} - \alpha}, & \textit{если } \alpha > 1 - p_n; \end{cases} \tag{30}$$

- *Если заданы распределение вероятностей* $Q = \{q_i, i \in \mathcal{N}\}$, $q_1 \geqslant q_2 \geqslant \ldots \geqslant q_n > 0$, *и число* $\alpha$, $0 \leqslant \alpha \leqslant 1$, *то*

$$\chi^2_{\max}(Q, \alpha) = \begin{cases} \dfrac{(1+\alpha-q_n)^2}{q_n} + \dfrac{(q_n-\alpha)^2}{q_{n-1}} - 1, & \textit{если } \alpha \leqslant q_n, \\ \dfrac{(1-\alpha)^2}{q_n} + \dfrac{(1-\alpha)^2}{q_{n-1}}, & \textit{если } \alpha \geqslant 1 - q_n; \end{cases} \tag{31}$$

- *Для всех* $\alpha$, $q_n < \alpha < 1 - q_n$, *справедлива верхняя граница*

$$\chi^2_{\max}(Q, \alpha) \leqslant \frac{(1-\alpha)^2}{q_n} + 1 - \alpha, \tag{32}$$

*причем эта верхняя граница достигается, т.е. в* (32) *имеет место знак равенства, если* $\alpha = q_n + \sum\limits_{i=1}^{n-1} a_i q_i$ *при некоторых* $a_i \in \{0, 1\}$.

Из этой теоремы также можно вывести приведенное ниже следствие (подобное следствию 1) для величин $\chi^2_{\max}(p_{\min}, \alpha)$ и $\chi^2_{\max}(q_{\min}, \alpha)$, определяемых равенствами

$$\chi^2_{\max}(p_{\min}, \alpha) = \max_{(P,Q):\, s(P,Q)=\alpha,\ \min\limits_{i\in\mathcal{N}} p_i = p_{\min}} \chi^2(P \,\|\, Q), \tag{33}$$

$$\chi^2_{\max}(q_{\min}, \alpha) = \max_{(P,Q):\, s(P,Q)=\alpha,\ \min\limits_{i\in\mathcal{N}} q_i = q_{\min}} \chi^2(P \,\|\, Q), \tag{34}$$

где максимумы в (33), (34) берутся по всевозможным распределениям $P = \{p_i, i \in \mathcal{N}\}$ и $Q = \{q_i,\ i \in \mathcal{N}\}$ с заданными параметрами $p_{\min} > 0$ в (33) и $q_{\min} > 0$ в (34), таким что $s(P, Q) = \alpha$.

С л е д с т в и е 2. *Для величин* $\chi^2_{\max}(p_{\min}, \alpha)$ *и* $\chi^2_{\max}(q_{\min}, \alpha)$, *определенных в* (33) *и* (34), *в случае* $|\mathcal{N}| = n \geqslant 3$ *справедливы следующие утверждения*:

- *Для всех* $p_{\min} > 0$ *и* $\alpha$, $0 \leqslant \alpha \leqslant 1$, *справедливо равенство*

$$\chi^2_{\max}(p_{\min}, \alpha) = \begin{cases} \infty, & \textit{если } \alpha \leqslant 1 - p_{\min}, \\ \dfrac{2p_{\min}(1-\alpha)^2}{p^2_{\min} - (1-\alpha)^2}, & \textit{если } \alpha > 1 - p_{\min}; \end{cases} \tag{35}$$

- *Для всех* $q_{\min} > 0$ *справедливо равенство*

$$\chi^2_{\max}(q_{\min}, \alpha) = \begin{cases} \dfrac{(1+\alpha-q_{\min})^2 + (q_{\min}-\alpha)^2}{q_{\min}} - 1, & \text{если } \alpha \leqslant q_{\min}, \\ \dfrac{2(1-\alpha)^2}{q_{\min}}, & \text{если } \alpha \geqslant 1 - q_{\min}; \end{cases} \tag{36}$$

- *Для всех* $q_{\min} > 0$ *и* $\alpha$, $q_{\min} < \alpha < 1 - q_{\min}$, *справедлива верхняя граница*

$$\chi^2_{\max}(q_{\min}, \alpha) \leqslant \frac{(1-\alpha)^2}{q_{\min}} + 1 - \alpha, \tag{37}$$

*причем эта верхняя граница достигается, т.е. в* (37) *имеет место знак равенства, если* $2q_{\min} \leqslant \alpha < 1 - q_{\min}$ *и* $q_{\min} \leqslant \dfrac{1}{n+1}$, *а также если* $q_{\min} \leqslant \dfrac{1}{n}$ *и* $\alpha = kq_{\min}$, *где* $k$ *– любое целое, такое что* $2 \leqslant k \leqslant n-1$.

Доказательство этого следствия вполне аналогично приведенному выше доказательству следствия 1 и поэтому здесь не приводится.

## § 2. Доказательства

Доказательство теоремы 1. Доказательство равенств (5) и (6) проводится вполне аналогично. Поэтому докажем, например, первое из них. Для доказательства того, что $D_f^{\min}(P, \alpha) = 0$ при $\alpha \geqslant 2p_{\max} - 1$ воспользуемся следующим утверждением (см. [3, теорема 1]): если $P = \{p_i,\ i \in \mathcal{N}\}$ и $Q = \{q_i,\ i \in \mathcal{N}\}$ – два распределения вероятностей, а $\alpha$, $0 \leqslant \alpha \leqslant 1$, – некоторое действительное число, то $\alpha$-склеивание $P$ и $Q$ (т.е. такое, что $s(P,Q) = \alpha$) существует тогда и только тогда, когда

$$\max_{i \in \mathcal{N}} [p_i + q_i - 1]^+ \leqslant \alpha \leqslant \sum_{i \in \mathcal{N}} \min\{p_i, q_i\}, \quad \text{где } [x]^+ = \begin{cases} x, & \text{если } x \geqslant 0, \\ 0, & \text{если } x \leqslant 0. \end{cases}$$

Из этого сразу следует, что $D_f^{\min}(P, \alpha) = 0$, если $p_{\max} \leqslant 1/2 + \alpha/2$, так как в этом случае существует $\alpha$-склеивание распределения $P$ с собой, а $D_f(P \,\|\, P) = 0$, так как по предположению $f(1) = 0$.

Поэтому надо доказать лишь второе из равенств в (5), когда предполагается, что $0 \leqslant \alpha \leqslant 2p_{\max} - 1$. Для этого вначале заметим, что для любых распределений вероятностей $P = \{p_i,\ i \in \mathcal{N}\}$ и $Q = \{q_i,\ i \in \mathcal{N}\}$ справедливо неравенство

$$D_f(P \,\|\, Q) \geqslant d_f(p_i \,\|\, q_i) \quad \text{для любых } i \in \mathcal{N}, \tag{38}$$

где $d_f(\cdot \,\|\, \cdot)$ определено в (4). Действительно, пользуясь свойством выпуклости функции $f(t)$, имеем

$$D_f(P \,\|\, Q) = q_i f\left(\frac{p_i}{q_i}\right) + \sum_{j\colon j \neq i} q_j f\left(\frac{p_j}{q_j}\right) =$$
$$= q_i f\left(\frac{p_i}{q_i}\right) + (1-q_i) \sum_{j\colon j \neq i} \frac{q_j}{1-q_i} f\left(\frac{p_j}{q_j}\right) \geqslant$$
$$\geqslant q_i f\left(\frac{p_i}{q_i}\right) + (1-q_i) f\left(\sum_{j\colon j \neq i} \frac{p_j}{1-q_i}\right) = d_f(p_i \,\|\, q_i).$$

Предположим теперь, что $Q = \{q_i,\ i \in \mathcal{N}\}$ – некоторое распределение вероятностей, для которого при заданном $\alpha$, $0 \leqslant \alpha \leqslant 2p_{\max} - 1$, существует его $\alpha$-склеивание с распределением $P = \{p_i,\ i \in \mathcal{N}\}$, у которого, для определенности, $p_{\max} = p_1$. Пусть также матрица $M = \|p_{ij}\|_{i,j=1}^n$ задает совместное распределение, осуществляющее это $\alpha$-склеивание $P$ и $Q$, т.е. $\sum\limits_{j=1}^n p_{ij} = p_i$ для всех $i \in \mathcal{N}$, $\sum\limits_{i=1}^n p_{ij} = q_j$ для всех $j \in \mathcal{N}$ и $\sum\limits_{i=1}^n p_{ii} = \alpha$. Тогда, воспользовавшись неравенством (38), получаем

$$D_f(P \,\|\, Q) \geqslant d_f(p_1 \,\|\, q_1) \geqslant d_f(p_{\max} \,\|\, 1 - p_{\max} + \alpha). \tag{39}$$

Второе неравенство в (39) следует из того, что функция $d_f(p_1 \,\|\, q_1)$ убывает по $q_1$ при $q_1 \leqslant p_1$, а в нашем случае $q_1 \leqslant \alpha + 1 - p_1 \leqslant p_1$, так как

$$q_1 \leqslant p_{11} + \sum_{i=2}^n p_i \leqslant \alpha + 1 - p_1 \leqslant p_1,$$

поскольку $0 \leqslant \alpha \leqslant 2p_1 - 1$.

Поэтому для доказательства второго равенства в (5) достаточно найти распределение вероятностей $Q = \{q_i,\ i \in \mathcal{N}\}$, для которого существует его $\alpha$-склеивание (при $0 \leqslant \alpha \leqslant 2p_{\max} - 1$) с заданным распределением $P = \{p_i,\ i \in \mathcal{N}\}$, и такое что

$$D_f(P \,\|\, Q) = d_f(p_{\max} \,\|\, 1 - p_{\max} + \alpha).$$

Действительно, легко убедиться, что такое распределение $Q = \{q_i,\ i \in \mathcal{N}\}$ является маргинальным для совместного распределения $P$ и $Q$, осуществляющего их $\alpha$-склеивание и задаваемого матрицей $M = \|p_{ij}\|_{i,j=1}^n$ с компонентами

$$p_{ij} = \begin{cases} \alpha & \text{при } i = j = 1, \\ cp_j & \text{при } i = 1 \text{ и } j \in \mathcal{N} \setminus \{1\}, \\ p_i & \text{при } i \in \mathcal{N} \setminus \{1\} \text{ и } j = 1, \\ 0 & \text{в остальных случаях,} \end{cases}$$

где параметр $c = \dfrac{p_1 - \alpha}{1 - p_1}$. ▲

Доказательство теоремы 2. Прежде всего заметим, что в рассматриваемом случае, когда задано распределение вероятностей $P = \{p_i,\ i \in \mathcal{N}\}$ и предполагается, что $\min\limits_i p_i = p_n > 0$, то из определения (1) следует, что для любого распределения $Q = \{q_i,\ i \in \mathcal{N}\}$, которое имеет хотя бы одно $q_i = 0$, справедливо равенство

$$D_f(P \,\|\, Q) = \sum_{i:\, q_i > 0} q_i f\left(\frac{p_i}{q_i}\right) + f^* \sum_{i:\, q_i = 0} p_i, \tag{40}$$

где $f^* = \lim\limits_{t \to \infty} \dfrac{f(t)}{t}$. Поэтому, если существует матрица $M = \|p_{ij}\|_{i,j=1}^n$, осуществляющая $\alpha$-склеивание распределения $P$ с некоторым распределением $Q$, у которой имеется столбец, целиком состоящий из нулей (т.е. некоторое $q_i = 0$), а $f^* = \infty$, то в этом случае $D_f^{\max}(P, \alpha) = \infty$. Очевидно, что такая матрица всегда существует, если $\alpha \leqslant 1 - p_n$, т.е. в этом случае справедливо первое равенство в (11). Поэтому в дальнейшем при доказательстве теоремы 1 всегда будет предполагаться, что либо $f^* < \infty$, либо, если $f^* = \infty$, то $\alpha > 1 - p_n$.

Для доказательства формулы (10) нужно показать, что существует *оптимальная матрица* $M = \|p_{ij}\|_{i,j=1}^n$ (т.е. матрица, для которой $D_f(M) = D_f^{\max}(P, \alpha)$), осуществляющая $\alpha$-склеивание заданного распределения вероятностей $P = \{p_i,\ i \in \mathcal{N}\}$

с некоторым распределением $Q = \{q_i,\ i \in \mathcal{N}\}$ и принадлежащая множеству $\mathcal{M}(P, I)$ для некоторого допустимого $(P, I)$-представления заданного числа $\alpha$.

В дальнейшем, для краткости, когда речь идет о некоторой матрице, будем всегда считать, что эта матрица осуществляет $\alpha$-склеивание заданного распределения вероятностей $P = \{p_i,\ i \in \mathcal{N}\}$ с некоторым распределением $Q = \{q_i,\ i \in \mathcal{N}\}$. Докажем вначале, что существует оптимальная матрица, у которой все ненулевые элементы (за исключением, возможно, лишь одного) расположены в некотором (главном) столбце и на (главной) диагонали.

Пусть $k$-й столбец матрицы $M = \|p_{ij}\|_{i,j=1}^n$ таков, что $\dfrac{q_k}{p_k} = \max\limits_{i \in \mathcal{N}} \dfrac{q_i}{p_i}$. Покажем, что $D_f(M)$ можно увеличить, если к каждому элементу (кроме диагонального) этого $k$-го столбца прибавить все элементы соответствующей строки, кроме диагонального. Действительно, для этого достаточно доказать, что $D_f(M(x)) > D_f(M)$, где $M(x) = \|p_{ij}(x)\|_{i,j=1}^n$ – матрица с элементами

$$p_{ij}(x) = \begin{cases} p_{\ell\, k} + x & \text{при } i = \ell \text{ и } j = k, \\ p_{\ell\, m} - x & \text{при } i = \ell \text{ и } j = m, \\ p_{ij} & \text{в остальных случаях,} \end{cases}$$

где $0 < x \leqslant p_{\ell\, m}$, а $\ell$ и $m$ – любые индексы, такие что $\ell \neq m$, $\ell \neq k$ и $m \neq k$. Имеем

$$[D_f(M(x)]'_x = [f(u) - uf'(u)] - [f(v) - vf'(v)],$$

где

$$u = \frac{p_k}{q_k + x}, \quad v = \frac{p_m}{q_m - x}.$$

Замечая теперь, что $u < v$, так как $x > 0$, а $\dfrac{p_k}{q_k} \leqslant \dfrac{p_m}{q_m}$ по условию, мы видим, что $f(u) - uf'(u)$ убывает по $u$ (так как $f(\cdot)$ – выпуклая функция), а поэтому $D_f(M(x))$ возрастает по $x$, и значит, $D_f(M(x)) > D_f(M(0)) = D_f(M)$.

Аналогично доказывается, что $D_f(M)$ можно увеличить, если все элементы $k$-й строки (когда $\dfrac{q_k}{p_k} = \max\limits_{i \in \mathcal{N}} \dfrac{q_i}{p_i}$ ), кроме диагонального, прибавить к одному из них. Очевидно, что без ограничения общности всегда можно считать, что если $k$-й столбец матрицы $M = \|p_{ij}\|_{i,j=1}^n$ является главным, то $\dfrac{q_k}{p_k} = \max\limits_{i \in \mathcal{N}} \dfrac{q_i}{p_i}$.

Чтобы доказать, что существует оптимальная матрица, принадлежащая некоторому множеству $\mathcal{M}(P, I)$, остается лишь показать, что существует оптимальная матрица, у которой на диагонали стоят некоторые числа $p_i$, $i \in \mathcal{N}$, а также, возможно, одно число $\beta$, $0 < \beta < p_j$, для некоторого $j \in \mathcal{N}$, и нули (последнее возможно, лишь если $\alpha \leqslant 1 - p_n$). Для этого достаточно доказать, что любая матрица, у которой на диагонали стоят по крайней мере два элемента, отличные от нуля и некоторых $p_i$, $i \in \mathcal{N}$, не может быть оптимальной.

Действительно, пусть $M = \|p_{ij}\|_{i,j=1}^n$ – некоторая матрица, у которой на диагонали стоят два элемента $p_{\ell\ell}$ и $p_{mm}$, $\ell \neq m$, такие что $0 < p_{\ell\ell} < p_\ell$ и $0 < p_{mm} < p_m$. Покажем, что в этом случае существует другая матрица $M(x)$, такая что $D_f(M(x)) > > D_f(M)$. Для этого необходимо рассмотреть два различных случая: когда ни $p_{\ell\ell}$, ни $p_{mm}$ не принадлежат главному столбцу матрицы $M$ и когда либо $p_{\ell\ell}$, либо $p_{mm}$ принадлежат ему. Оба случая анализируются вполне аналогично. Поэтому рассмотрим, например, первый из них, когда в матрице $M$ главным столбцом является $k$-й, а $\ell \neq k$ и $m \neq k$. Пусть для определенности $\dfrac{q_\ell}{p_\ell} \geqslant \dfrac{q_m}{p_m}$. В этом случае рассмотрим матрицу $M(x) = \|p_{ij}(x)\|_{i,j=1}^n$ с элементами

$$p_{ij}(x) = \begin{cases} p_{\ell\ell} + x & \text{при } i = j = \ell, \\ p_{mm} - x & \text{при } i = j = m, \\ p_{\ell k} - x & \text{при } i = \ell \text{ и } j = k, \\ p_{mk} + x & \text{при } i = m \text{ и } j = k, \\ p_{ij} & \text{в остальных случаях,} \end{cases}$$

где $x > 0$ достаточно мало. Тогда, очевидно, имеем

$$\frac{q_\ell(x)}{p_\ell(x)} = \frac{q_\ell + x}{p_\ell} > \frac{q_m(x)}{p_m(x)} = \frac{q_m - x}{p_m} \quad \text{и} \quad \frac{q_i(x)}{p_i(x)} = \frac{q_i}{p_i} \quad \text{для всех } i \neq \ell \text{ и } i \neq m.$$

Поэтому, как мы видели выше, $D_f(M(x))$ возрастает по $x$, и значит, $D_f(M(x)) > D_f(M)$. Таким образом, мы доказали, что существует оптимальная матрица, принадлежащая множеству $\mathcal{M}(P, I)$ при некотором допустимом $(P, I)$-представлении числа $\alpha$. Отсюда сразу следует справедливость формулы (10).

Для доказательства второго из равенств в (11) заметим, что в данном случае предполагается, что компоненты $p_i$ распределения $P = \{p_i,\ i \in \mathcal{N}\}$ упорядочены по убыванию и $\alpha > 1 - p_n$. Поэтому из формулы (10) сразу следует, что оптимальную матрицу следует искать среди матриц $M = \|p_{ij}\|_{i,j=1}^n$, у которых на диагонали стоят числа $p_i$, $i \in \mathcal{N} \setminus \{k\}$, и $p_k - (1 - \alpha)$ при некотором $k$, вне диагонали в некотором $j$-м $(j \neq k)$ столбце стоит число $1 - \alpha$, а все остальные элементы матрицы равны нулю. Для такой матрицы

$$D_f(M) = (p_k + \alpha - 1) f\left(\frac{p_k}{p_k + \alpha - 1}\right) + (1 - \alpha + p_j) f\left(\frac{p_j}{1 - \alpha + p_j}\right).$$

Замечая теперь, что, как нетрудно убедиться, функции

$$(p_k + \alpha - 1) f\left(\frac{p_k}{p_k + \alpha - 1}\right) \quad \text{и} \quad (1 - \alpha + p_j) f\left(\frac{p_j}{1 - \alpha + p_j}\right)$$

убывают по $p_k$ и $p_j$ соответственно (здесь существенно, что в соответствии с определением $f$-дивергенции выпуклая функция $f(\cdot)$ такова, что $f(1) = 0$), мы видим, что для максимума $D_f(M)$ среди подобных матриц $M$, т.е. для $D_f^{\max}(P, \alpha)$, справедливо второе из равенств (11), где $K_P$ определено в (12). ▲

Отметим, что хотя в общем случае нельзя сказать, какое из двух выражений в определении $K_P$ является максимальным, однако во многих частных случаях $f$-дивергенции это можно сделать. В частности, это удается сделать для дивергенции Кульбака–Лейблера и $\chi^2$-дивергенции (см. доказательства теорем 4 и 5 ниже).

Доказательство теоремы 3. Прежде всего заметим, что в рассматриваемом случае, когда задано распределение вероятностей $Q = \{q_i,\ i \in \mathcal{N}\}$ и предполагается, что $\min\limits_i q_i = q_n > 0$, то вместо (40) из определения (1) следует, что для любого распределения $P = \{p_i,\ i \in \mathcal{N}\}$, которое имеет хотя бы одно $p_i = 0$, имеет место равенство

$$D_f(P \,\|\, Q) = \sum_{i:\, p_i > 0} q_i f\left(\frac{p_i}{q_i}\right) + f(0) \sum_{i:\, p_i = 0} q_i, \tag{41}$$

а тогда снова очевидно (как и при доказательстве теоремы 2), что справедливо первое из равенств в (14). Дальнейшее доказательство этой теоремы вполне аналогично приведенному выше доказательству теоремы 2 и поэтому здесь не приводится. ▲

Доказательство теоремы 4. **1.** Докажем вначале равенство (20). Так как дивергенция Кульбака–Лейблера $D(P\,\|\,Q)$ является $f$-дивергенцией при $f(t) = t\log t$, то равенство (20) является следствием соотношения (11), поскольку в данном случае $f^* = \infty$, а величина $K_P$ (см. (12)), как нетрудно убедиться, равна

$$p_n \log \frac{p_n}{p_n - 1 + \alpha} + p_{n-1} \log \frac{p_{n-1}}{p_{n-1} + 1 - \alpha}.$$

Действительно, для этого следует лишь заметить, что разность первого и второго выражений в (12) при $f(t) = t\log t$ убывает по $\alpha$, а при $\alpha = 1$ она равна нулю.

**2.** Докажем теперь равенства (21) и (22). Из общей формулы (13) теоремы 3 следует, что в случае, когда $\alpha \leqslant q_n$, существует оптимальная матрица (осуществляющая $\alpha$-склеивание заданного распределения вероятностей $Q = \{q_i,\ i \in \mathcal{N}\}$ с некоторым распределением $P = \{p_i,\ i \in \mathcal{N}\}$), находящаяся среди матриц

$$M_1(k,\ell) = \big\|p_{ij}^{(1)}\big\|_{i,j=1}^{n}, \quad M_2(k,\ell) = \big\|p_{ij}^{(2)}\big\|_{i,j=1}^{n}, \quad M_3(k,\ell,m) = \big\|p_{ij}^{(3)}\big\|_{i,j=1}^{n}$$

(где $k$, $\ell$ и $m$ – всевозможные различные между собой числа, принадлежащие множеству $\mathcal{N}$) с элементами

$$p_{ij}^{(1)} = \begin{cases} \alpha & \text{при } i = j = k, \\ q_i & \text{при } i \in \mathcal{N} \setminus \{k\} \text{ и } j = k, \\ q_k - \alpha & \text{при } i = k \text{ и } j = \ell, \\ 0 & \text{в остальных случаях,} \end{cases} \tag{42}$$

$$p_{ij}^{(2)} = \begin{cases} q_i & \text{при } i \in \mathcal{N} \setminus \{k,\ell\} \text{ и } j = k, \\ q_\ell - \alpha & \text{при } i = \ell \text{ и } j = k, \\ q_k & \text{при } i = k \text{ и } j = \ell, \\ \alpha & \text{при } i = j = \ell, \\ 0 & \text{в остальных случаях,} \end{cases} \tag{43}$$

$$p_{ij}^{(3)} = \begin{cases} q_i & \text{при } i \in \mathcal{N} \setminus \{k,m\} \text{ и } j = k, \\ q_m - \alpha & \text{при } i = m \text{ и } j = k, \\ q_k & \text{при } i = k \text{ и } j = \ell, \\ \alpha & \text{при } i = j = m, \\ 0 & \text{в остальных случаях.} \end{cases} \tag{44}$$

Заметим, что в каждой из этих матриц $k$-й столбец является главным. Таким образом, имеем

$$D^{\max}(Q,\alpha) = \max_{k,\ell,m}\{D(M_1(k,\ell)), D(M_2(k,\ell)), D(M_3(k,\ell,m))\},$$

где

$$D(M_1(k,\ell)) = (1 + \alpha - q_k)\log\frac{1 + \alpha - q_k}{q_k} + (q_k - \alpha)\log\frac{q_k - \alpha}{q_\ell},$$

$$D(M_2(k,\ell)) = (1 - \alpha - q_k)\log\frac{1 - \alpha - q_k}{q_k} + (q_k + \alpha)\log\frac{q_k + \alpha}{q_\ell},$$

$$D(M_3(k,\ell,m)) = (1 - \alpha - q_k)\log\frac{1 - \alpha - q_k}{q_k} + q_k\log\frac{q_k}{q_\ell} + \alpha\log\frac{\alpha}{q_m}.$$

Очевидно, что

$$\max_{k,\ell,m} D(M_3(k,\ell,m)) \leqslant \max_{k,\ell} D(M_2(k,\ell)),$$

и поскольку $D(M_1(k,\ell))$ и $D(M_2(k,\ell))$ убывают по $q_\ell$ и являются выпуклыми функциями от $q_k$, то

$$D^{\max}(Q,\alpha)=\max\{D(M_i(n-1,n)),D(M_i(1,n)),D(M_i(n,n-1)),\ i=1,2\}.$$

Легко видеть, что

$$D(M_1(1,n))\leqslant\max\{D(M_1(n-1,n)),D(M_2(n,n-1))\},$$
$$D(M_2(1,n))\leqslant\max\{D(M_2(n-1,n)),D(M_1(n,n-1))\},$$

а поэтому для доказательства равенства (21) (правая часть которого совпадает с выражением для $D(M_1(n,n-1))$ ) необходимо доказать, что каждая из трех величин $D(M_1(n-1,n))$, $D(M_2(n-1,n))$ и $D(M_2(n,n-1))$ не превосходит $D(M_1(n,n-1))$.

Неравенство $D(M_1(n,n-1))\geqslant D(M_1(n-1,n))$ является следствием того, что разность $D(M_1(n,n-1))-D(M_1(n-1,n))$, как легко проверить, возрастает по $\alpha$, а при $\alpha=0$ она положительна. Действительно, нетрудно убедиться, что эта последняя разность $[D(M_1(n,n-1))-D(M_1(n-1,n))]\big|_{\alpha=0}$ является убывающей функцией $q_n$, а при максимальном значении $q_n$, равном $q_{n-1}$, она равна нулю.

Доказательство двух оставшихся неравенств $D(M_1(n,n-1))\geqslant D(M_2(n-1,n))$ и $D(M_1(n,n-1))\geqslant D(M_2(n,n-1))$ проводится вполне аналогично. Таким образом, равенство (21) доказано.

Справедливость равенства (22) легко следует из соотношения (14) теоремы 3, так как в рассматриваемом случае, когда $f(t)=t\log t$, величина $K_Q$ (см. (15)) равна

$$(q_n-1+\alpha)\log\frac{q_n-1+\alpha}{q_n}+(1-\alpha+q_{n-1})\log\frac{1-\alpha+q_{n-1}}{q_{n-1}}.$$

Действительно, нетрудно убедиться, что разность первого и второго выражений в (15) убывает по $\alpha$, а при $\alpha=1$ она равна нулю.

**3.** Докажем теперь верхнюю границу (23). Для этого необходимо рассмотреть три типа матриц $M=\|p_{ij}\|_{i,j=1}^n$, принадлежащих одному из множеств $\mathcal{M}(Q,I)$, когда:

а) в главном столбце матрицы на диагонали стоит некоторое $q_k$, $k\in I$, входящее в одно из допустимых $(Q,I)$-представлений $\alpha$ в виде $\alpha=\sum\limits_{i\in I}q_i+\beta$;

б) в главном столбце матрицы на диагонали стоит ноль;

в) в главном столбце матрицы на диагонали стоит число $\beta$, входящее в одно из допустимых $(Q,I)$-представлений $\alpha$ в виде $\alpha=\sum\limits_{i\in I}q_i+\beta$.

Согласно общей формуле (13) среди таких матриц находится оптимальная, и нам надо показать, что для всех таких матриц $M$ справедлива верхняя граница

$$D(M)\leqslant(1-\alpha+q_n)\log\frac{1-\alpha+q_n}{q_n}$$

при всех $\alpha$, $q_n\leqslant\alpha\leqslant1-q_n$. Рассмотрим вкратце доказательства этой границы для каждого из этих трех типов матриц.

а) Если в главном столбце матрицы на диагонали стоит некоторое $q_k$, $k\in I$, входящее в одно из допустимых $(Q,I)$-представлений $\alpha$ в виде $\alpha=\sum\limits_{i\in I}q_i+\beta$, а $\beta$ стоит на диагонали в $j$-м столбце, то для такой матрицы

$$D(M)=(1-\alpha+q_k)\log\frac{1-\alpha+q_k}{q_k}+\beta\log\frac{\beta}{q_j}\leqslant(1-\alpha+q_n)\log\frac{1-\alpha+q_n}{q_n}. \qquad (45)$$

Действительно, неравенство в этом соотношении следует из того, что функция

$$(1-\alpha+q_k)\log\frac{1-\alpha+q_k}{q_k}$$

убывает по $q_k$, а $\beta \log \dfrac{\beta}{q_j} \leqslant 0$, так как $0 \leqslant \beta < q_j$ по условию $(Q, I)$-допустимого представления $\alpha$. Заметим сразу, что в (45) вместо неравенства справедливо равенство (а значит, и равенство в формуле (23)), если $k = n$ и $\beta = 0$, т.е. если существует $(Q, I)$-представление $\alpha$ вида $\alpha = q_n + \sum\limits_{i=1}^{n-1} a_i q_i$ при некоторых $a_i \in \{0, 1\}$.

б) Если в главном ($k$-м) столбце матрицы $M$ на диагонали стоит ноль, а число $\beta$, входящее в одно из допустимых $(Q, I)$-представлений $\alpha$, стоит на диагонали в $j$-м столбце, то нетрудно видеть, что при любом расположении элемента $q_k$ в матрице $M$ величину $D(M)$ можно оценить сверху следующим образом:

$$D(M) \leqslant \max_{(k,j):\, k \neq j} \left\{ (1 - \alpha - q_k) \log \frac{1 - \alpha - q_k}{q_k} + (q_k + q_j) \log \frac{q_k + q_j}{q_j} \right\}. \tag{46}$$

Полагая

$$g(\alpha, q_k, q_j) = (1 - \alpha - q_k) \log \frac{1 - \alpha - q_k}{q_k} + (q_k + q_j) \log \frac{q_k + q_j}{q_j},$$

заметим, что разность

$$(1 - \alpha + q_n) \log \frac{1 - \alpha + q_n}{q_n} - g(\alpha, q_k, q_j)$$

является убывающей функцией $\alpha$ при любых $q_k$ и $q_j$, а при максимальном значении $\alpha$ (если фиксировано любое $q_k$), равном $1 - q_k$, имеем

$$\left[ (1 - \alpha + q_n) \log \frac{1 - \alpha + q_n}{q_n} - g(\alpha, q_k, q_j) \right] \Bigg|_{\alpha = 1 - q_k} =$$
$$= (q_n + q_k) \log \frac{q_n + q_k}{q_n} - (q_j + q_k) \log \frac{q_j + q_k}{q_j} \geqslant 0,$$

так как $(q_j + q_k) \log \dfrac{q_j + q_k}{q_j}$ является убывающей функцией $q_j$. Поэтому из (46) следует, что для рассматриваемого класса матриц $M$ (у которых в главном столбце на диагонали стоит ноль) также справедливо доказываемое неравенство

$$D(M) \leqslant (1 - \alpha + q_n) \log \frac{1 - \alpha + q_n}{q_n}.$$

в) Наконец, если в главном ($k$-м) столбце матрицы $M$ на диагонали стоит число $\beta$, входящее в одно из допустимых $(Q, I)$-представлений $\alpha$, то снова нетрудно видеть, что при любом расположении элемента $q_k - \beta$ в матрице $M$ величину $D(M)$ можно оценить сверху следующим образом:

$$D(M) \leqslant \max_{(k,j):\, k \neq j} \left\{ (1 - \alpha - q_k + 2\beta) \log \frac{1 - \alpha - q_k + 2\beta}{q_k} + \right.$$
$$\left. + (q_k + q_j - \beta) \log \frac{q_k + q_j - \beta}{q_j} \right\}. \tag{47}$$

Замечая теперь, что максимизируемая в правой части (47) функция является выпуклой относительно $\beta$, приходим к выводу, что максимум правой части (47) достигается либо при $\beta = 0$, либо при $\beta = q_k$, что сводит задачу к рассмотренным выше случаям а) и б). На этом доказательство верхней границы (23) заканчивается. ▲

Доказательство теоремы 5. **1.** Равенство (30) и второе из равенств в (31) (где $\alpha \geqslant 1 - q_n$) являются прямыми следствиями формул (11), (12) и (14), (15) для рассматриваемого случая $f(t) = (t-1)^2$.

**2.** Доказательство первого из равенств в (31) (где $\alpha \leqslant q_n$) в основном следует схеме доказательства формулы (21) теоремы 4, и поэтому мы опишем его лишь вкратце.

Оптимальную матрицу, осуществляющую $\alpha$-склеивание заданного распределения вероятностей $Q = \{q_i,\ i \in \mathcal{N}\}$ с некоторым распределением $P = \{p_i,\ i \in \mathcal{N}\}$, снова следует искать среди матриц

$$M_1(k,\ell) = \left\|p_{ij}^{(1)}\right\|_{i,j=1}^{n}, \quad M_2(k,\ell) = \left\|p_{ij}^{(2)}\right\|_{i,j=1}^{n}, \quad M_3(k,\ell,m) = \left\|p_{ij}^{(3)}\right\|_{i,j=1}^{n}$$

с элементами, заданными равенствами (42)–(44), которые определяют $\chi^2$-дивергенции следующих трех типов:

$$\chi^2(M_1(k,\ell)) = \frac{(1+\alpha-q_k)^2}{q_k} + \frac{(q_k-\alpha)^2}{q_\ell} - 1,$$
$$\chi^2(M_2(k,\ell)) = \frac{(1-\alpha-q_k)^2}{q_k} + \frac{(q_k+\alpha)^2}{q_\ell} - 1,$$
$$\chi^2(M_3(k,\ell,m)) = \frac{(1-\alpha-q_k)^2}{q_k} + \frac{q_k^2}{q_\ell} + \frac{\alpha^2}{q_m} - 1,$$

где $k$, $\ell$ и $m$ – всевозможные различные между собой числа, принадлежащие множеству $\mathcal{N}$. Так как очевидно, что

$$\max_{k,\ell,m} \chi^2(M_3(k,\ell,m)) \leqslant \max_{k,\ell} \chi^2(M_2(k,\ell)),$$

а $\chi^2(M_1(k,\ell))$ и $\chi^2(M_2(k,\ell))$ являются выпуклыми функциями относительно $q_k$ и убывающими относительно $q_\ell$, то рассуждения, приведенные при доказательстве в пункте 2 теоремы 4, здесь полностью сохраняются (с соответствующей заменой величин $D(M_i(\cdot\,,\cdot))$ на $\chi^2(M_i(\cdot\,,\cdot))$ ) и показывают, что в рассматриваемом случае $\alpha \leqslant q_n$ имеем

$$\begin{aligned}&\chi^2_{\max}(Q,\alpha) = \max\bigl\{\chi^2(M_i(n,n-1)), \chi^2(M_i(n-1,n)),\ i=1,2\bigr\} = \\ &= \chi^2(M_1(n,n-1)),\end{aligned}$$

что и доказывает первое из равенств в (31).

**3.** Доказательство верхней границы (32) также в основном следует схеме доказательства верхней границы (23) в теореме 4. Снова нам необходимо доказать, что для трех типов матриц $M = \|p_{ij}\|_{i,j=1}^{n}$, принадлежащих одному из множеств $\mathcal{M}(Q,I)$ и введенных в пункте 3 доказательства теоремы 4, справедлива верхняя граница (32), т.е. для всех таких матриц $M$ справедливо неравенство

$$\chi^2(M) \leqslant \frac{(1-\alpha)^2}{q_n} + 1 - \alpha,$$

если $q_n \leqslant \alpha \leqslant 1 - q_n$. Докажем это утверждение.

а) Если в главном столбце матрицы $M$ на диагонали стоит некоторое $q_k$, $k \in I$, входящее в одно из допустимых $(Q,I)$-представлений $\alpha$ в виде $\alpha = \sum\limits_{i\in I} q_i + \beta$, а $\beta$ стоит на диагонали в $j$-м столбце, то для такой матрицы

$$\chi^2(M) = \sum_{i=1}^{n} \frac{p_i^2}{q_i} - 1 = \frac{(1-\alpha+q_k)^2}{q_k} + \frac{\beta^2}{q_j} + \sum_{i:\, p_i=q_i} q_i - 1 =$$
$$= \frac{(1-\alpha)^2}{q_k} + 1 - \alpha + \frac{\beta^2}{q_j} - \beta \leqslant \frac{(1-\alpha)^2}{q_n} + 1 - \alpha,$$

так как $0 \leqslant \beta < q_j$ по условию $(Q, I)$-допустимого представления $\alpha$. Снова замечаем, что вместо вышеприведенного неравенства имеет место равенство, если $\beta = 0$ и $k = n$, т.е. если существует $(Q, I)$-представление $\alpha$ вида $\alpha = q_n + \sum\limits_{i=1}^{n-1} a_i q_i$ при любых $a_i \in \{0, 1\}$.

б) Если в главном ($k$-м) столбце матрицы $M$ на диагонали стоит ноль, а число $\beta$, входящее в одно из допустимых $(Q, I)$-представлений $\alpha$, стоит на диагонали в $j$-м столбце, то нетрудно видеть, что при любом расположении элемента $q_k$ в матрице $M$ величину $\chi^2(M)$ можно оценить сверху следующим образом:

$$\chi^2(M) \leqslant \max_{(k,j):\, k\neq j} \left\{ \frac{(1-\alpha)^2}{q_k} + \frac{q_k^2}{q_j} - 3(1-\alpha-q_k) \right\} \leqslant \frac{(1-\alpha)^2}{q_n} + 1 - \alpha. \qquad (48)$$

Действительно, справедливость второго неравенства в (48) следует из того, что разность

$$\frac{(1-\alpha)^2}{q_n} + 1 - \alpha - \frac{(1-\alpha)^2}{q_k} - \frac{q_k^2}{q_j} + 3(1-\alpha-q_k),$$

как легко убедиться, является убывающей функцией $\alpha$ при любых $q_k$ и $q_j$, а при максимальном значении $\alpha = 1 - q_k$ (так как на диагонали в главном $k$-м столбце матрицы стоит ноль) эта разность неотрицательна.

в) Если в главном ($k$-м) столбце матрицы $M$ на диагонали стоит число $\beta$, входящее в одно из допустимых $(Q, I)$-представлений $\alpha$, то снова нетрудно видеть, что при любом расположении элемента $q_k - \beta$ в матрице $M$ величину $\chi^2(M)$ можно оценить сверху следующим образом:

$$\chi^2(M) \leqslant \max_{(k,j):\, k\neq j} \left\{ \frac{(1-\alpha-q_k+2\beta)^2}{q_k} + \frac{(q_k+q_j-\beta)^2}{q_j} + \alpha - \beta - q_j - 1 \right\}. \qquad (49)$$

Поскольку максимизируемая в правой части (49) функция является выпуклой относительно $\beta$, то максимум правой части (49) достигается либо при $\beta = 0$, либо при $\beta = q_k$, что, как нетрудно убедиться, сводит задачу к рассмотренным выше случаям а) и б). ▲

## СПИСОК ЛИТЕРАТУРЫ


1. *Csiszár I.* Information-type Measures of Difference of Probability Distributions and Indirect Observations // Studia Sci. Math. Hungar. 1967. V. 2. P. 299–318.
2. *Ali S.M., Silvey S.D.* A General Class of Coefficients of Divergence of One Distribution from Another // J. Roy. Statist. Soc. Ser. B. 1966. V. 28. № 1. P. 131–142. https://www.jstor.org/stable/2984279
3. *Sason I., Verdú S.* $f$-Divergence Inequalities // IEEE Trans. Inform. Theory. 2016. V. 62. № 11. P. 5973–6006. https://doi.org/10.1109/TIT.2016.2603151
4. *Макур А., Чжен Л.* Сравнение коэффициентов сжатия для $f$-дивергенций // Пробл. передачи информ. 2020. Т. 56. № 2. С. 3–62. https://doi.org/10.1134/S0134347520020011

5. *Прелов В.В.* О склеивании вероятностных распределений и оценивании дивергенции через вариацию // Пробл. передачи информ. 2017. Т. 53. № 3. С. 16–22. http://mi.mathnet.ru/ppi2239
6. *Прелов В.В.* О некоторых оптимизационных задачах для дивергенции Реньи // Пробл. передачи информ. 2018. Т. 54. № 3. С. 36–53. http://mi.mathnet.ru/ppi2271
7. *Gilardoni G.L.* On the Minimum $f$-Divergence for Given Total Variation // C. R. Math. Acad. Sci. Paris. 2006. V. 343. № 11–12. P. 763–766. https://doi.org/10.1016/j.crma.2006.10.027
8. *Gilardoni G.L.* On Pinsker's and Vajda's Type Inequalities for Csiszár's $f$-Divergences // IEEE Trans. Inform. Theory. 2010. V. 56. № 11. P. 5377–5386. https://doi.org/10.1109/TIT.2010.2068710
9. *Прелов В.В.* О максимальных значениях $f$-дивергенции и дивергенции Реньи при заданном вариационном расстоянии // Пробл. передачи информ. 2020. Т. 56. № 1. С. 3–15. https://doi.org/10.1134/S0134347520010015

*Прелов Вячеслав Валерьевич*
Институт проблем передачи информации
им. А.А. Харкевича РАН
prelov@iitp.ru

Поступила в редакцию
17.11.2020
После доработки
04.01.2021
Принята к публикации
11.01.2021